

○○
○○○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Classification automatique de flux radiophoniques par Machines à Vecteurs de Support

Mathieu Ramona

Directeurs de Thèse : Gaël Richard, Bertrand David

RTL - TELECOM ParisTech

21 juin 2010





Vers une radio numérique

- L'un des derniers médias analogiques.
- La numérisation implique des services supplémentaires.
- Traitement automatisé des archives.
- Quelles méta-données ?



Classe acoustique

Identification préalable de la classe acoustique :

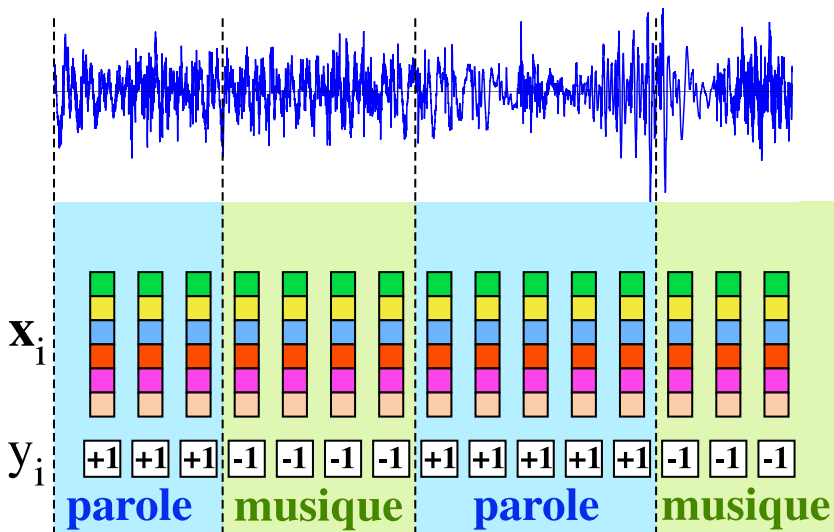
- Parole
- Musique
- Chant
- ...

○○○
 ○○○○○○○○○○
 ○○○○○○○○

○○○
 ○○○
 ○○○○○

○○○○○
 ○○
 ○○

Classification audio





Caractérisation audio

Descripteurs dans la littérature

Apports très variés :

Parole, Spectraux, Banc de filtres, Perceptuels...

Lesquels exploiter ?

- Ensemble très vaste.
- Nécessité d'une sélection :
 - Complexité de calcul
 - Redondances
 - Présence possible de bruit.
 - Compréhension du problème.
- Sélection manuelle sub-optimale.
- Comment définir la pertinence d'un descripteur ?
- Quel lien avec le classifieur ?

⇒ Algorithmes de **sélection automatique de descripteurs**.

○○○
○○○○○○○○○○○
○○○○○○○○○

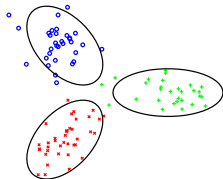
○○○
○○○
○○○○○

○○○○○
○○
○○

État de l'art - Méthodes de classification

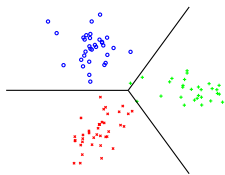
Méthodes génératives

- Modèle gaussien
- Mélanges de gaussiennes
- Modèles de Markov Cachés



Méthodes discriminatives

- Heuristiques
- Analyse Discriminante Linéaire
- k Plus Proches Voisins
- Réseaux de Neurones
- **Machines à Vecteurs de Support**



Approches hybrides

- SVM/GMM, Supervecteurs

○○
○○○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Machines à Vecteurs de Support (SVM)

Essor des SVM dans le domaine de l'apprentissage statistique.

Avantages

- Discriminatifs
- Maximisation de la Marge
- Sur-apprentissage contrôlé
- *Kernel trick*
- Polyvalence du noyau
- Peu de paramètres

Contraintes

- Bi-classe
- Complexité
- Hyper-paramètres



Problématiques

- **Optimisation des SVM** : comment fixer efficacement les paramètres du noyau, de manière optimale ?
- **Multi-classes** : comment appliquer les SVM sur un problèmes impliquant plus de deux classes ?
- **Sélection de descripteurs** : comment introduire un cadre efficace de sélection des descripteurs, adapté au cadre du modèle SVM ?
- **Aspect temporel** : comment introduire la continuité temporelle du flux audio dans une approche par sac de trames ?
- **Traitement en ligne** : contraintes liée à la complexité, la garantie d'un traitement *en ligne* est souhaitée.

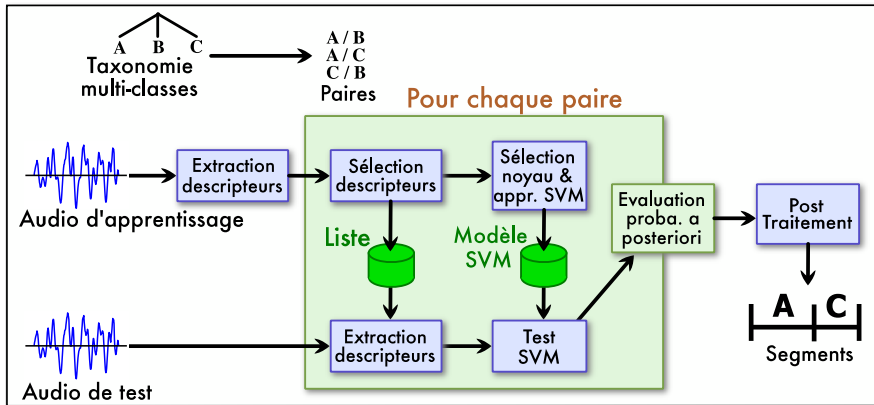
L'accent mis sur les problèmes liés à l'application des SVM sur une tâche pratique de classification audio.

○○○
○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Systeme global



○○○
○○○○○○○○○○○○○
○○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

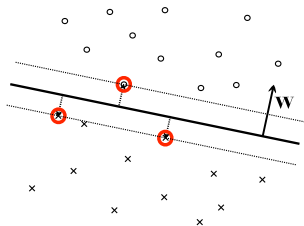
Plan

- 1 Introduction
- 2 Machines à Vecteurs de Support
 - Théorie
 - Sélection de noyau
 - Sélection de descripteurs
- 3 SVM pour la classification audio
 - SVM multi-classes
 - Approches dynamiques
 - Segmentation aveugle
- 4 Évaluations
 - ESTER 1
 - ESTER 2
 - Démonstration
- 5 Conclusion



Principe des Machines à Vecteurs de Support

- **Données** : $\mathcal{S} = \{(\mathbf{x}_i, y_i)_{i=1\dots n}\}$
avec $y_i = \begin{cases} +1 & \text{pour } 1 \leq i \leq n_1 \\ -1 & \text{pour } n_1 < i \leq n \end{cases}$
- **Problème** : prédiction des y_i .
- **Séparation linéaire** : $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$
- Critère : minimiser la **marge** $M = \min \frac{y_i f(\mathbf{x}_i)}{\|\mathbf{w}\|}$.



Problèmes d'optimisation quadratique

	Primal	Dual	$([H]_{ij} = y_i y_j \mathbf{x}_i^T \mathbf{x}_j)$
Objectif	$\min_{\mathbf{w}} \frac{1}{2} \ \mathbf{w}\ ^2$	$\max_{\alpha} L_D(\alpha) = \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T \mathbf{H} \alpha$	
Contraintes	$y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0$	$0 \leq \alpha_i, \quad \alpha^T \mathbf{y} = 0$	

Vecteurs de Supports

- Vecteur normal : $\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$
- Seuls comptent $\alpha_i \neq 0 \Rightarrow$ **vecteurs de supports**.
- \Rightarrow **Eparsification** du problème.



Noyaux de Mercer

Kernelisation

- \mathcal{X} : espace de descripteurs.
- $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ est un *noyau semi-défini positif* si :

$$\int_{\mathcal{X}} k(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad \forall g \in \mathcal{C}(\mathcal{X}).$$
- **Théorème de Mercer** : $\exists \Phi / k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$
- **Kernel trick** : k remplace tout produit scalaire, sans exprimer Φ .
- $\dim \mathcal{F} > \dim \mathcal{X}$
- Hyperplan linéaire dans $\mathcal{F} \Rightarrow$ surface plus complexe dans \mathcal{X} .
- **SVM kernelisés** : même problème, avec $[\mathbf{H}]_{ij} = y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$.

Noyaux usuels

- Noyau polynômial : $k(\mathbf{x}, \mathbf{y}) = (1 + c \mathbf{x}^T \mathbf{y})^\delta$
- Noyau RBF gaussien : $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$
- Noyau sigmoïdal : $k(\mathbf{x}, \mathbf{y}) = \tanh(c \mathbf{x}^T \mathbf{y} + \theta)$



Hyper-paramètres

Les noyaux non-linéaires incluent un ou plusieurs *hyper-paramètres* :

- Noyau polynômial : c et δ .
- Noyau RBF gaussien : σ .
- Noyau sigmoïdal : c et θ .

Illustration

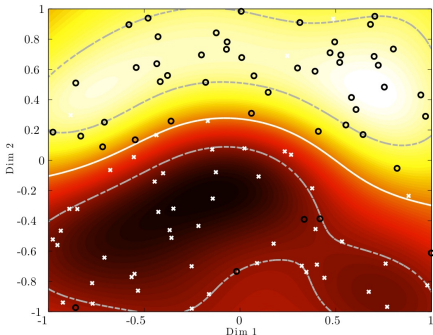
Distribution avec *outliers*.

Paramètre σ du noyau RBF gaussien :

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$$

Agit sur la distance entre les exemples :

- σ trop petit : sur-apprentissage.
- σ trop grand : exemples indifférenciés.
- σ optimal.



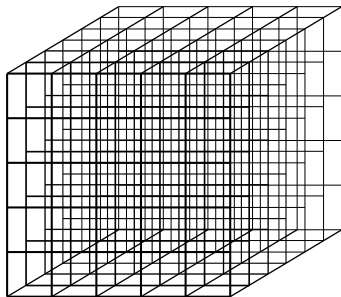
$\sigma = 0.7$



Détermination des hyper-paramètres

Approche classique

- Minimiser l'erreur par **validation croisée** sur l'ensemble d'apprentissage.
- Stratégie de **recherche par maillage**.
- Plusieurs défauts :
 - Complexité polynomiale \Rightarrow explosion combinatoire.
 - Champ de recherche limité par le maillage.
 - Réduction de l'ensemble d'apprentissage



216 valeurs

Approche par optimisation

\Rightarrow Introduction de critères **dérivables** d'évaluation du noyau.



Alignement du noyau (KTA)

Définition

- Exemples $\mathcal{S} = \{(\mathbf{x}_i, y_i)_{i=1\dots n}\}$, avec $y_i = \begin{cases} +1 & \text{pour } 1 \leq i \leq n_1 \\ -1 & \text{pour } n_1 < i \leq n \end{cases}$
- Noyau k
- Matrice de Gram : $[\mathbf{K}]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$
- Matrice Cible (*Target*) : $\mathbf{K}^* = \mathbf{y}\mathbf{y}^T$.
- Blocs de classes : $\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix} \quad \mathbf{K}^* = \begin{pmatrix} \mathbf{1} & -\mathbf{1} \\ -\mathbf{1} & \mathbf{1} \end{pmatrix}$.
- Produit de Frobenius : $\langle \mathbf{A}, \mathbf{B} \rangle_F = \sum_{i,j=1}^n a_{ij}b_{ij}$
- L'**Alignement du noyau** (ou KTA : Kernel Target Alignment)
est le produit de Frobenius normalisé de \mathbf{K} et de la matrice cible \mathbf{K}^* :

$$\mathcal{A}(\mathbf{K}, \mathbf{K}^*) = \frac{\langle \mathbf{K}, \mathbf{K}^* \rangle_F}{\|\mathbf{K}\|_F \|\mathbf{K}^*\|_F}$$

(nous appellerons $\mathcal{F}(\mathbf{K}, \mathbf{K}^*) = \langle \mathbf{K}, \mathbf{K}^* \rangle_F$ le **critère de Frobenius**)



Interprétation de l'alignement

Interprétation géométrique

- Développement du kernel trick : $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$
- On montre que

$$\langle \mathbf{K}, \hat{\mathbf{K}}^* \rangle_F = \left(\frac{1}{n_1} \sum_{\mathbf{x}_i \in \mathcal{S}_1} \phi(\mathbf{x}_i) - \frac{1}{n_2} \sum_{\mathbf{x}_j \in \mathcal{S}_2} \phi(\mathbf{x}_j) \right)^2 = \|\boldsymbol{\mu}_1^\Phi - \boldsymbol{\mu}_2^\Phi\|^2$$

- Critère de Frobenius = **dispersion inter-classes** dans l'espace \mathcal{F} .

Fenêtres de Parzen

- Estimateur de densité de probabilité : $\hat{f}_n(\mathbf{x}) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$
- Fenêtre gaussienne : $K_G(\|\mathbf{x} - \mathbf{y}\|/\sigma) = (2\pi)^{-1/2} k_{\text{rbf}}(\mathbf{x}, \mathbf{y})$
- Mesure de séparabilité : $\mathcal{P}_{\text{sep}}(\mathbf{x}_j, \mathbf{y}_j) = \hat{f}(\mathbf{x}_j | \mathbf{y} = \mathbf{y}_j) - \hat{f}(\mathbf{x}_j | \mathbf{y} \neq \mathbf{y}_j)$

- On montre : $\mathcal{P}_{\text{sep}}(\mathcal{S}) = \frac{1}{\sqrt{2\pi}\sigma} \langle \mathbf{K}, \hat{\mathbf{K}}^* \rangle_F$

- Critère de Frobenius = **mesure de séparabilité** dans l'espace \mathcal{X} .



Séparabilité de Classes Kernelisée (KCS)

- Critère de Frobenius = variance inter-classes dans \mathcal{F} .

- Matrice de **dispersion inter-classes** :

$$\mathbf{S}_b = \frac{1}{n} \sum_{c=1,2} n_c (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^T$$

- Matrice de **dispersion intra-classes** :

$$\mathbf{S}_w = \sum_{c=1,2} \sum_{\mathbf{x}_i \in \mathcal{S}_c} (\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T$$

- Critère de **Séparabilité des classes** : $\mathcal{C} = \frac{\text{tr } \mathbf{S}_b}{\text{tr } \mathbf{S}_w}$

- *Kernelisation* du critère : $\mathcal{C}_K = \frac{\mathbf{1}_n^T \mathbf{B} \mathbf{1}_n}{\mathbf{1}_n^T \mathbf{W} \mathbf{1}_n}$

où \mathbf{B} et \mathbf{W} s'expriment à partir de la matrice de Gram \mathbf{K} .

- **Critère KCS** : mesure de séparabilité des classes dans l'espace \mathcal{F} .



Algorithme de sélection

Dérivation

- Soit un noyau k_{Θ} de paramètres $\Theta = [\theta_1, \dots, \theta_P]$.
- La dérivation est immédiate :

$$\frac{\partial}{\partial \theta_p} \mathcal{A}(\mathbf{K}_{\Theta}, \mathbf{K}^*) = \frac{\langle \partial_{\theta_p} \mathbf{K}_{\Theta}, \mathbf{K}^* \rangle_F}{\|\mathbf{K}_{\Theta}\|_F \|\mathbf{K}^*\|_F} - \frac{\langle \mathbf{K}_{\Theta}, \mathbf{K}^* \rangle_F \langle \mathbf{K}_{\Theta}, \partial_{\theta_p} \mathbf{K}_{\Theta} \rangle_F}{\|\mathbf{K}_{\Theta}\|_F^3 \|\mathbf{K}^*\|_F}$$

où l'on introduit $\partial_{\theta_p} \mathbf{K}_{\Theta} = [\partial_{\theta_p} k_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)]_{ij}$.

- Optimisation du noyau k_{Θ} par **descente de gradient** :
 - Maximisation du KTA ou du KCS.
 - Mise à jour des paramètres θ_p .

Régularisation

- **Convergence triviale** du KCS ($\mathcal{C}_K \rightarrow 1$) avec le noyau RBF gaussien.

- Ajout d'un terme de régularisation : $\mathcal{C}'_K = \frac{\mathbf{1}_n^T \mathbf{B} \mathbf{1}_n}{\mathbf{1}_n^T \mathbf{W} \mathbf{1}_n + n^2 \epsilon}$



Machines à Marge Souple

Variables d'écart

- On a jusque là supposé la **séparabilité** des données.
- Relâchement des contraintes par les **variables d'écart** $\xi_i > 0$:

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i$$

- Exemple \mathbf{x}_i **hors-marge** ssi $\xi_i > 1$.

Marge souple L1 et L2

- SVM à marge souple : $\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i^k$.
- Le **facteur d'erreur** C fixe le compromis entre les deux critères.

- L1-SVM** ($k = 1$) : $\max_{\alpha} \quad \alpha^T \mathbf{1} - \frac{1}{2} \mathbf{a}^T \mathbf{K} \mathbf{a} \quad (\mathbf{a} = \alpha \bullet \mathbf{y})$
avec $\alpha^T \mathbf{y} = 0$ et $0 \leq \alpha_i \leq C$

- L2-SVM** ($k = 2$) : $\max_{\alpha} \quad \alpha^T \mathbf{1} - \frac{1}{2} \mathbf{a}^T (\mathbf{K} + \frac{1}{C} \mathbf{I}) \mathbf{a}$
avec $\alpha^T \mathbf{y} = 0$ et $0 \leq \alpha_i$



Optimisation du paramètre C

C hyper-paramètre ?

- C n'est pas un paramètre du noyau.
 - ⇒ absent de la matrice de Gram.
 - ⇒ ignoré par les critères KTA et KCS.
- L1 SVM généralement employés mais...
- L2 SVM = marge dure avec $\mathbf{K}_C = \mathbf{K} + \frac{1}{C} \mathbf{I}$
 - ⇒ calcul de l'Alignement sur \mathbf{K}_C .
- Coût négligeable :

$$\begin{aligned} \langle \mathbf{K}_C, \mathbf{K}^* \rangle_F &= \langle \mathbf{K}, \mathbf{K}^* \rangle_F + \frac{n}{C} \\ \|\mathbf{K}_C\|_F^2 &= \|\mathbf{K}\|_F^2 + \frac{n}{C^2} \\ \Sigma(\mathbf{B}_C) &= \Sigma(\mathbf{B}) - \frac{n-2}{C} \\ \Sigma(\mathbf{W}_C) &= \Sigma(\mathbf{W}) + \frac{n-2}{C} \end{aligned}$$

Quelle valeur pour C ?

Valeur de Joachims (*SVMLight* [113]) : $C_{\text{def}} = \frac{1}{\overline{R^2}}$



Validation expérimentale

Protocole

- Recherche du σ optimal sur un noyau RBF gaussien.
- Calcul des différents critères sur une grille logarithmique entre 0.1 et 20.
- Comparaison à l'erreur **Leave One Out** sur l'optimum de chaque critère.
- 4 bases de tests : *Spambase*, *Ionosphere*, *Lymphoma*, *Parole/Musique*.

Etat de l'art

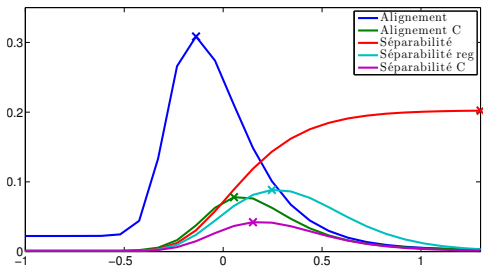
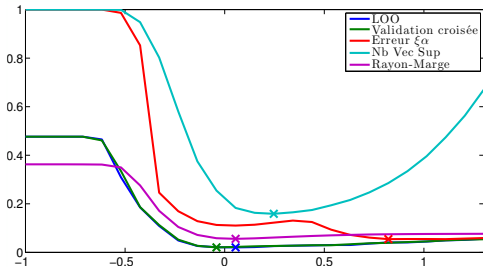
- Validation croisée (10 fold).
- Taux de Vecteurs de Supports.
- Estimée $\xi\alpha$ (Joachims 2000 [115]).
- Borne Rayon-Marge (Vapnik 98 [230]).

○○○
 ○○○○○○○○○○●
 ○○○○○○○○

○○○
 ○○○
 ○○○○○

○○○○○
 ○○
 ○○

Résultats sur Parole/Musique





Sélection de descripteurs

Principe

- Sélection des D_S descripteurs les plus pertinents parmi $D > D_S$.
- Notion de **pertinence** complexe :
 - **Interpertinence** entre descripteurs.
 - **Redondances** \Rightarrow sélection conjointe.
 - **Liée** au processus de classification.
- Approche classique : $R_{Fisher}(d) = \frac{|\mu_{1,d} - \mu_{2,d}|^2}{\sigma_{1,d}^2 + \sigma_{2,d}^2}$.
- Indépendante du classifieur.

Contributions

- Proposition de 4 algorithmes liés au SVM : SAS, FAS, SCSS, KFDS.
- Nous ne détaillerons ici que SAS et FAS.



Sélection de descripteurs (2)

Noyau pondéré

- **Définition** : $k_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = k(\mathbf{w} \bullet \mathbf{x}, \mathbf{w} \bullet \mathbf{y})$
avec les poids $\mathbf{w} = [w_1, \dots, w_D]$.
- $\Theta = \{w_1, \dots, w_D\}$ sont les paramètres du noyau
⇒ sélection de noyau.
- Optimisation pour déterminer les valeurs optimales w_d .
- w_d : pondération du descripteur d dans le noyau
⇒ mesure de **pertinence**
- Sélection de descripteurs = Sélection de noyau

Remarque : Optimisation implicite du paramètre σ :

$$k_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\sum_d w_d^2 (x_d - y_d)^2}{2\sigma^2}\right)$$

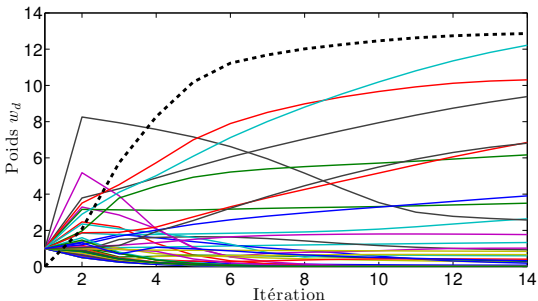


Scaled Alignment Selection (SAS)

- Maximisation de l'Alignement par rapport aux poids w_d .
- Optimisation par simple **descente de gradient**.
- Initialisation à $w_d = 1 \forall d$
- Dérivation :

$$\frac{\partial}{\partial w_d} \mathcal{A}(\mathbf{K}_w, \mathbf{K}^*) = \frac{\langle \partial_{w_d} \mathbf{K}_w, \mathbf{K}^* \rangle_F}{\|\mathbf{K}_w\|_F \|\mathbf{K}^*\|_F} - \frac{\langle \mathbf{K}_w, \mathbf{K}^* \rangle_F \langle \mathbf{K}_w, \partial_{w_d} \mathbf{K}_w \rangle_F}{\|\mathbf{K}_w\|_F^3 \|\mathbf{K}^*\|_F}$$

avec $[\partial_{w_d} \mathbf{K}_w]_{ij} = \partial_{w_d} k_w(\mathbf{x}_i, \mathbf{x}_j)$.





Matrices de Gram dimensionnelles

- Décomposition opportune des noyaux :

Noyau linéaire

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} \\ = \sum_d \kappa^d(\mathbf{x}, \mathbf{y})$$

$$\kappa^d(\mathbf{x}, \mathbf{y}) = x_d \cdot y_d$$

Noyau RBF Gaussien

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right) \\ = \exp\left(-\frac{\sum_d (x_d - y_d)^2}{2\sigma^2}\right) \\ = \prod_d \kappa^d(\mathbf{x}, \mathbf{y})$$

$$\kappa^d(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{(x_d - y_d)^2}{2\sigma^2}\right)$$

- κ^d représente la **contribution** du descripteur d au noyau k .
- On définit ainsi les **matrices de Gram dimensionnelles** :

$$\boldsymbol{\kappa}^d = [\kappa^d(\mathbf{x}_i, \mathbf{x}_j)]_{ij}$$



Forward Alignment Selection (FAS)

- **Principe** : sélection itérative du descripteur maximisant l'Alignement.
- **Initialisation** : K matrice de Gram "neutre"
(linéaire : $\mathbf{0}$ / RBF gaussien : $\mathbf{1}$).
- Pour chaque descripteur d non sélectionné :
 - Évaluation de la matrice de Gram avec la matrice de Gram dimensionnelle :
linéaire : $K \leftarrow K + \kappa^d$ / RBF gaussien : $K \leftarrow K \bullet \kappa^d$
 - Calcul de l'Alignement
- **Mise à jour** de la matrice de Gram en incluant le descripteur sélectionné.



Validation expérimentale

Protocole expérimental

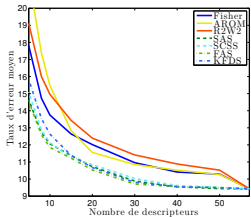
- Expérience sur différentes bases de données réelles (*Spambase*, *Ionosphere* et *Parole/Musique*)
- Extraction d'un ensemble d'apprentissage et de test.
- Sélection de D_S descripteurs sur l'ensemble d'apprentissage.
- Apprentissage d'un SVM sur les D_S descripteurs.
- Comparaison sur noyaux linéaire et RBF gaussien.
- Calcul de l'erreur sur l'ensemble de test (moyenne sur 30 itérations).

INCOMPLET.

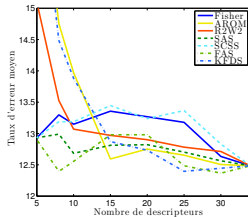


Résultats

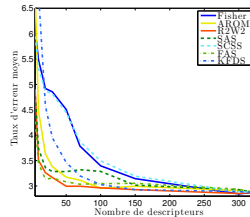
Linéaire



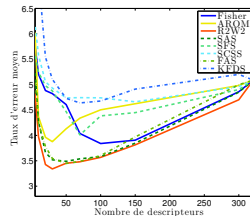
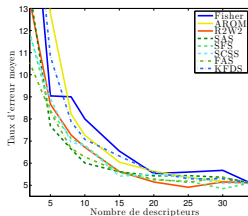
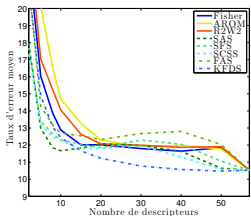
Ionosphere



Parole/Musique



RBF gaussien



INCOMPLET.



Problème multi-classes

- **Problème RTL** : parole/musique/parole+musique/chant.
- SVM : modèle pour $C = 2$ classes.
- \Rightarrow quelles solutions $C > 2$?

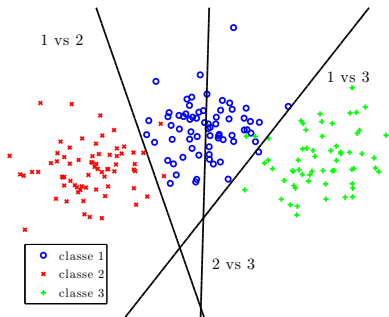
Approches classiques

Un contre tous (OVA) :

- Un classifieur f_c pour chaque classe c contre les autres.
 $\Rightarrow C$ classifieurs.
- Décision : $\arg \max_c f_c(\mathbf{x})$.

Un contre un (OVO) :

- Un classifieur $f_{c,d}$ pour chaque paire de classes c et d .
 $\Rightarrow \frac{C(C-1)}{2}$ classifieurs.
- Décision : PWC (Hastie & Tibshirani).

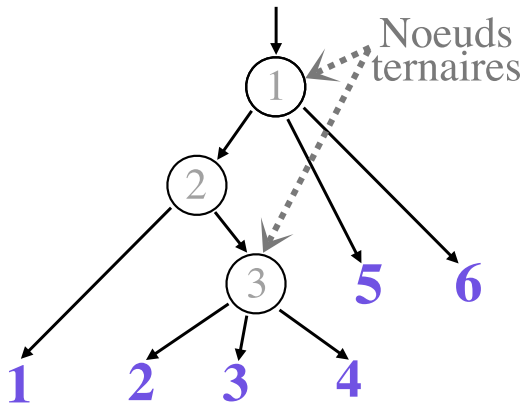


○○
 ○○○○○○○○○○
 ○○○○○○

○○●
 ○○○
 ○○○○

○○○○○
 ○○
 ○○

Approche hiérarchique



Estimation des **probabilités a posteriori**
 + **hybridation** avec nœuds multi-classes Un-contre-Un



Principe du post-traitement

Importance des probabilités a posteriori

- Continuité temporelle
- Représentation compacte

Lissage par filtrage

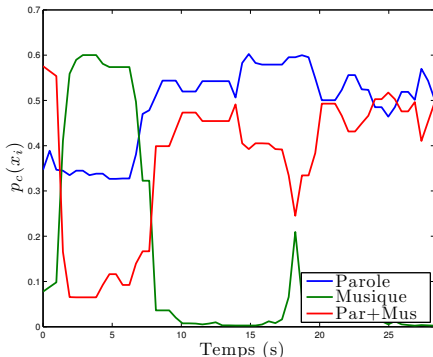
Sur fenêtres glissantes.

Extrait audio : bulletin d'informations.

- **Filtrage moyen**
lissage des transitions.
- **Filtrage médian**
meilleures transitions.

⇒ Horizon local (compromis).

⇒ Classes non distinguées.





Post-traitement par HMM

Problème considéré :

- A partir d'**observations** ... (les prob. a posteriori $\mathbf{p}(i) = [p_1(i), \dots, p_C(i)]$)
- ... déduire des **états** (les labels de classe $y_i \in [1, \dots, C]$).
- \Rightarrow Modèles de Markov a états Cachés.

Modèle HMM

- *Prob. d'observation* $b_c(\mathbf{p}(i))$: Modèle GMM à M composantes.
- *Prob. d'état initial* π_c : annotations.
- *Prob. de transitions* a_{cd} : tramage des annotations.

Modèle appris sur un ensemble de validation. Post-traitement : **Algorithme de Viterbi**.

Modèles Semi-Markoviens

Probabilité de *stagnation* : $e_c(L)$

- HMM impliquent pour $e_c(L)$ une distribution géométrique.
- HSMM : modélisation de $e_c(L)$.

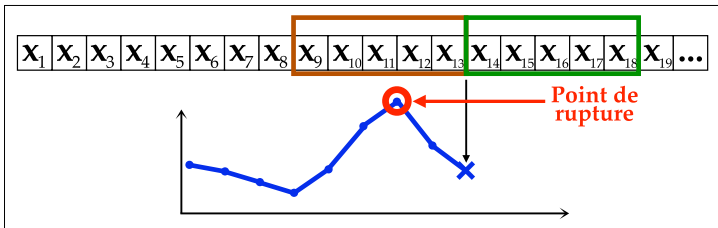


Segmentation aveugle

Principe

- Détection de rupture dans le signal audio.
- Origine : segmentation en locuteurs.
- Découpage en segments de classe acoustique homogène.
- À partir des descripteurs.

Fenêtres glissantes



⇒ Quelle métriques pour la détection de rupture ?

⇒ Quelle stratégie pour la recherche de maxima ?



Détection de rupture (1)

Critère d'Information Bayésienne (BIC)

- 3 modèles : P_1 passé, P_2 futur, P_0 sur les deux fenêtres.
- **Test de vraisemblance** entre les deux hypothèses :
 - H_0 : pas de rupture, modèle global P_0 .
 - H_1 : rupture observée, deux modèles P_1 et P_2 .
- + pénalisation de la **complexité**
 \Rightarrow modèle gaussien (μ, Σ) : $k = d + \frac{d(d+1)}{2}$.

Mesures probabilistes dans les espaces RKHS

- 2 mesures probabilistes sur des ensembles :
 - Divergence de Kullback-Leibler
 - Distance de Bhattacharyya
- **Kernelisation** des mesures.
- Modèles **gaussiens** dans l'espace \mathcal{F} .



Détection de rupture (2) - SVM à une classe

- Estimation du support d'une distribution \mathcal{S} :

$$f(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathcal{S}$$

$$f(\mathbf{x}) < 0 \quad \forall \mathbf{x} \notin \mathcal{S}.$$

- Même problèmes liés au sur-apprentissage :
compromis optimal
- \Rightarrow Minimisation du Risque Structurel.
- Fonction de décision : $f(\mathbf{x}) = \text{sign}(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho)$.

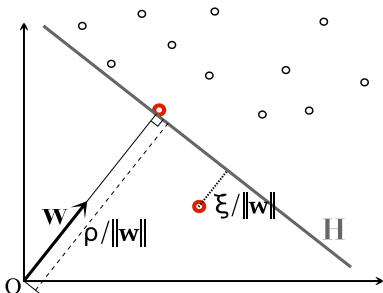
Rapport de vraisemblance (LLR)

Entre les hypothèses :

- Fenêtre W_2 générée par modèle P_1 .
- Fenêtre W_2 générée par modèle P_2 .

Kernel Change Detection (KCD)

Distance dans l'espace \mathcal{F} entre les modèles P_1 et P_2 .



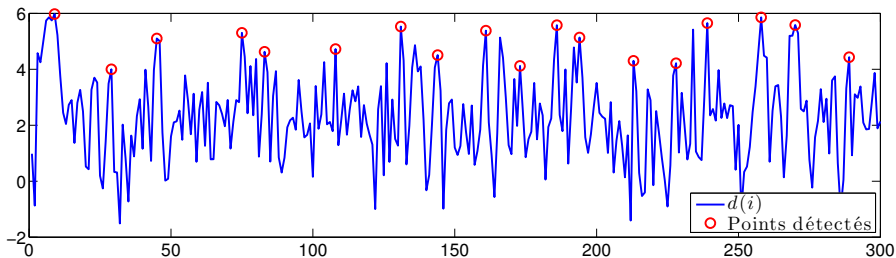
○○○
○○○○○○○○○○○○○
○○○○○○○○○

○○○
○○○
○○○●

○○○○○
○○○
○○

Recherche de maxima

- Dynamiques hétérogènes sur un long signal \Rightarrow harmonisation.
- Élagage des pics voisins \Rightarrow filtre max "plateau".
- Seuillage.





Évaluations sur ESTER 1

Campagne ESTER 1

- Campagne nationale d'évaluation de transcription.
- Tâche SES : détection de parole et de musique.
- Classes disjointes : parole/musique/mix.
- + réannotation : segments de chant.



Corpus ESTER

- Médias : généralistes (5 Radio France + RTM)
- Émissions : majorité bulletins d'informations.
- Apprentissage : 90h - Test : 10h

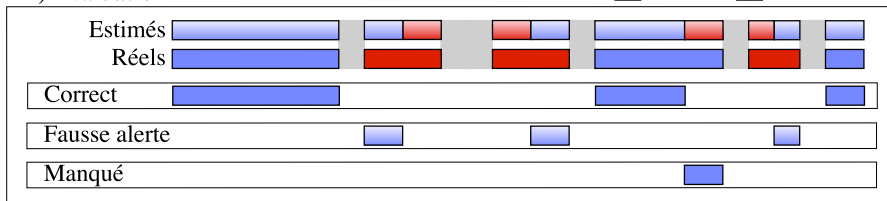
Ensemble	Chant	Mix	Musique	Parole+Br	Parole
Apprentissage	0.8%	10.1%	2.3%	6.0%	80.8%
Test	0.4%	12.5%	2.6%	5.2%	79.3%



Protocole d'évaluation

2) Evaluation

■ Classe ■ Non-classe



- Durées cumulées : d_{OK} , d_{FA} et d_{DM} .

- **Rappel** $R = \frac{d_{OK}}{d_{OK} + d_{DM}}$ **Précision** $P = \frac{d_{OK}}{d_{OK} + d_{FA}}$.

- **F-mesure** : $F = \frac{2RP}{R + P}$.

- **Fausse alerte** : $FA = \frac{d_{FA}}{d_{EST}}$ **Faux rejet** : $MD = \frac{d_{DM}}{d_{REEL}}$

avec $d_{EST} = d_{OK} + d_{FA}$
 $d_{REEL} = d_{OK} + d_{DM}$.

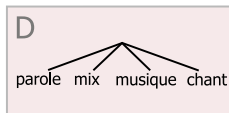
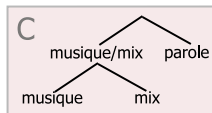
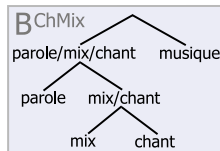
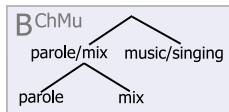
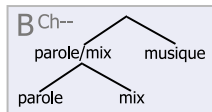
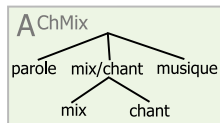
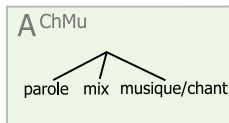
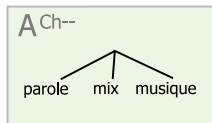
○○○
○○○○○○○○○○○○○
○○○○○○○○○

○○○
○○○
○○○○○

○○○●○○
○○
○○

Expérience : comparaison de taxonomies

- 2 taxonomies principales (A et B) + 2 annexes (C et D)
- Différentes modalités pour la prise en compte du chant.

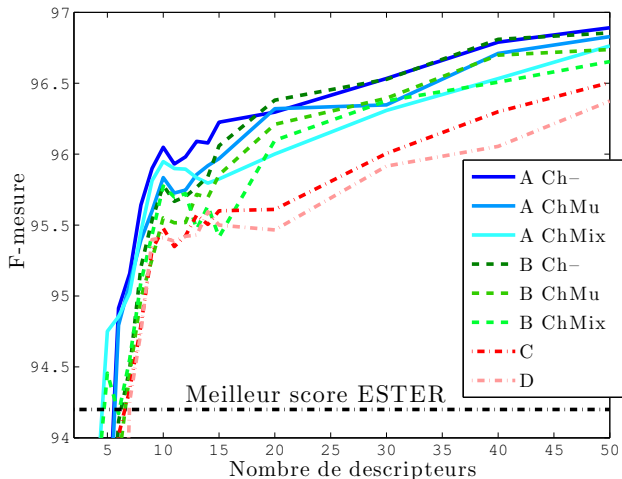


○○○
 ○○○○○○○○○○
 ○○○○○○○○

○○○
 ○○○
 ○○○○○

○○○○●○
 ○○
 ○○

Résultats (1)



- Nb de descripteurs variable.
- Sélection par IRMFSP.
- SVM à noyau RBF gaussien.
- 20000 exemples par classe.
- Filtrage médian sur 5 s.



Résultats (2)

Participant	globale			parole			musique		
	F	%fa	%fr	F	%fa	%fr	F	%fa	%fr
$d = 50$	96.9	2.0	4.5	99.4	13.0	0.5	78.8	1.5	29.6
$d = 10$	95.9	3.3	5.4	99.1	19.4	1.0	73.8	2.5	33.2
$d = 2$	93.3	11.9	4.1	98.9	16.2	1.5	64.8	11.6	20.3
1 ^{er} ESTER	94.2	2.1	9.5	98.8	30.1	1.5	52.7	1.2	61.7
2 ^e ESTER	93.1	1.3	12.1	98.9	9.7	1.9	33.7	1.0	78.5
3 ^e ESTER	92.7	11.7	5.7	99.2	36.6	0.7	54.8	10.9	38.7
4 ^e ESTER	90.7	1.3	16.2	97.4	8.0	4.9	17.8	1.1	89.6

Résultats avec la meilleurs taxonomie (A_{Ch-}) :

- Très bons résultats, même avec peu de descripteurs.
- Gain absolu global de 3.7% sur le meilleur résultats d'ESTER.
- Gain notable en détection de la musique (jusqu'à +26%), grâce à une réduction du taux de faux rejet.



Campagne ESTER 2

- Participation durant cette thèse.
- Tâche identique, corpus plus diversifié.

Système	Erreur(%)					md(%)	fa(%)	F
	Africa	Inter	RFI	TVME	Globale			
<i>Classe de parole</i>								
IRISA	1,65	1,42	0,58	2,44	1,49	0.37	16.42	99,20
IRIT	2,05	0,85	0,65	2,47	1,31	0.72	9.28	99,29
LIMSI	2,55	0,52	0,26	1,71	1,08	0.80	4.91	99,42
RTL	1,40	1,10	0,61	2,07	1,23	0.50	11.01	99,34
<i>Classe de musique</i>								
IRIT	6,63	5,17	5,93	4,63	5,51	43.13	0.77	69,80
TPT/RTL	12,40	2,95	3,92	4,10	5,25	12.56	4.33	78,85

○○
○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○
○○
○●

Démonstration

- Système de classification en temps réel.
- Implémentation en C++.

○○
○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Conclusion

- Proposition d'un **système efficace** tirant parti des SVM.
- Amélioration de critères existants pour la **sélection de noyau**.
- Proposition de nouveaux algorithmes de **sélection de descripteurs**.
- Comparaison de plusieurs méthodes de **post-traitement**.
- Très bons résultats aux **campagnes d'évaluation nationales**.
- Comparaison au niveau **international** au travers de corpora publics.
- **Implémentation** d'un système exploitable par l'entreprise.
- Mise en place d'un corpus public pour la **détection de chant**.

○○
○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Perspectives

- Problème de **régularisation du KCS**.
- Exploration des variantes aux SVM.
- Descripteur pour la détection de **double pitch**.
- Résultats décevants pour les **HSMM**.
- Prise en compte de la **structure sémantique**.

○○
○○○○○○○○○○○○
○○○○○○○○

○○○
○○○
○○○○○

○○○○○○
○○
○○

Merci pour votre attention...